

The Identity Issue

Susan Brown and Laura Mandell

02.13.18

Clusters: Gender

Article DOI: 10.22148/16.020

PDF: 10.31235/osf.io/9qvby

Journal ISSN: 2371-4549

Cite: Susan Brown and Laura Mandell, "The Identity Issue," Journal of Cultural Analytics. Feb. 13, 2018. DOI: 10.31235/osf.io/9qvby

This special issue of *Cultural Analytics* tackles the urgent question of how social identities can be addressed through computational methods. In particular, it probes the extent to which large datasets can be used to elucidate the kinds of questions that humanities scholars want to ask about historical and representational processes that structure social relations and positions. The papers to be published here emerge from the work of the Text Mining the Novel research partnership,¹ which chose identity as one of its themes of inquiry. This issue is thus situated at the intersection of literary studies and sociology, looking outward toward the novel's construction of ethnicities, genders, class categories, and racial terminology. It is also situated in an emerging scholarly publishing ecology that allows the journal to release articles progressively, as they pass through the peer review process and reach completion, rather than waiting for an entire set to be complete. This introduction therefore situates identity issues in relation to cultural analytics more broadly, and provides summaries of the provisional contents of the special issue.

¹The Text Mining the Novel project is led by Andrew Piper at McGill University and funded by a Partnership Grant from the Social Sciences and Humanities Research Council of Canada. See <http://www.noveltm.ca>.

Identity categories have played a role throughout the politics of modernity in movements associated with class, race, nationalism, religion, sexuality, and gender. There are significant continuities between these movements and post-World War II identity politics, not least in that the latter inherits the highly complex, historically shifting, and ideologically laden terminology for identities employed in the former.² What is commonly understood as identity politics, political movements grounded in self-identification and experiences of marginalization, emerged most clearly as a practice during the African-American civil rights agitation of the 1960s in the United States.³ The term itself emerged from the political activism of a range of groups, most notably of black lesbian feminists who became conscious of the impact of multiple forms of oppression,⁴ and was subsequently taken up in academic discourses from the 1980s, including within literary theory, in debates about “identity.”⁵

During the 1980s, cultural theorists expressed dissatisfaction with political activism in the name of an “identity” (woman, gay, Black, lesbian) insofar as it seemed to require “a belief in true essence [as] that which is most irreducible, unchanging, and therefore constitutive of a given person or thing.”⁶ At the same time, the mainstream feminist movement was under increasing pressure to account for the situatedness and complexity of gendered experience, particularly that of racialized or otherwise multiply marginalized subjects, who made it clear that there was no single unified group called “women.”⁷ From both contexts emerged a series of debates over the extent to which gender should be understood as socially constructed rather than essential.

This theoretical introduction focuses on gender identity for several reasons. First, some of the most fruitful theorizing of identity to date for the purposes of lit-

²Linda J. Nicholson, *Identity before Identity Politics* (Cambridge: Cambridge University Press, 2008).

³Craig Calhoun, *Social Theory and the Politics of Identity* (New York: Wiley, 1994).

⁴The coining of the phrase “identity politics” is attributed to the famous statement of the Combahee River Collective from 1977. Rpt. as “The Combahee River Collective Statement” in *Home Girls: A Black Feminist Anthology*, ed. Barbara Smith (New York: Kitchen Table/Women of Color Press, 1983), 264-274. See Michelle Murphy, *Seizing the Means of Reproduction: Entanglements of Feminism, Health, and Technoscience* (Durham, NC: Duke University Press, 2012), 39; and T. David Mason, “Ethnicity and Politics,” in *Encyclopedia of Government and Politics*, 2nd ed., 2 vols., ed. Mary E. Hawkesworth and Maurice Kogan (New York: Rutledge, 2004): 568-586.

⁵A comprehensive bibliography is available in Jodi Dean, *Solidarity of Strangers: Feminism After Identity Politics* (Berkeley: University of California Press, 1996); cited in Paula Moya, “Introduction,” *Reclaiming Identity: Realist Theory and the Predicament of Postmodernism*, ed. Paula Moya, Michael Hames-García (Berkeley: University of California Press, 2000), 1-26.

⁶Diana Fuss, *Essentially Speaking: Feminism, Nature, and Difference* (New York: Routledge, 1989), 2.

⁷See, for instance, Gloria T. Hull, Patricia Bell Scott, and Barbara Smith, *All the women are white, all the men are black, but some of us are brave* (New York: Feminist Press, 1982).

erary and cultural studies has been of gender identity, although it has been informed from the outset by perspectives, debates, and intellectual contributions emerging from the civil rights and the gay and lesbian liberation movements, and more recently by critical race studies, post-colonial, and post-human inquiry. Second, almost all of the work in quantitative computational approaches to identity in relation to literature has focused on gender, ranging from Gaye Tuchman and Nina E. Fortin's early sociological work on the publication of Victorian fiction (1989), through Shlomo Argamon and colleagues' use of vector analysis to classify French texts by the gender of their authors (2009), to Katherine Bode's ground-breaking revision of the history of Australian fiction, which takes up gender in a nuanced dialogue with nationalism and colonialism in her use of the AustLit database (2012), and Matt Jockers' mining of novels for gendered patterns in genre and discourse (2013).⁸ Finally, there has been a debate amongst feminist scholars in the digital humanities including Bethany Nowviskie, Lisa Marie Rhody, and Laura Mandell about the gendered discourse of text mining and its value as a mode of inquiry. This means that a consideration of gender as a category of analysis and its relation to methodology is inextricable from an evaluation of cultural analytics as a field.⁹

This introduction reviews some of the major interventions in the theory of feminist identity politics, highlighting as we go those ideas most relevant to investigating gender categories quantitatively, as a way of introducing the larger purpose of investigating identities by the articles gathered together in this issue. We take this deep dive into the history of gender theory because gender is both a central issue for some of the contributions to this cluster and is an illuminating case

⁸Gaye Tuchman and Nina Fortin, *Edging Women Out: Victorian Novelists, Publishers and Social Change* (New Haven: Yale University Press, 1989). Shlomo Argamon, Jean-Baptiste Goulain, Russell Horton, and Mark Olsen, "Vive la Différence! Text Mining Gender Difference in French Literature," *Digital Humanities Quarterly* 3.2 (2009). Katherine Bode, *Reading by Numbers: Recalibrating the Literary Field* (New York: Anthem Press, 2012). See also: Matt Jockers, *Macroanalysis* (Urbana: University of Illinois Press, 2013); Jan Rybicki, "Vive la différence: Tracing the (Authorial) Gender Signal by Multivariate Analysis of Word Frequencies," *Digital Scholarship in the Humanities* (2015): 1-16; Mark Olsen, "Écriture féminine: Searching for an Indefinable Practice?" *Literary and Linguistic Computing* 20 suppl. (2005): 147-64; Mark Olsen, "Qualitative Linguistics and *Histoire des Mentalités*: Gender Representation in the *Trésor de la Langue Française*," *QALICO* (14 January 1992); Andrew Piper and Richard Jean So, "Women Write About Family, Men Write About War," *The New Republic*, April 8, 2016.

⁹Bethany Nowviskie, "What Do Girls Dig?" *Debates in Digital Humanities*, ed. Matthew K. Gold (Minneapolis: University of Minnesota Press, 2012); Lisa Marie Rhody, "Why I Dig: Feminist Approaches to Text Analysis," in *Debates in Digital Humanities*, ed. Matthew K. Gold and Lauren Klein (Minneapolis: University of Minnesota Press, 2016); Laura Mandell, "Gender and Cultural Analytics: Finding or Making Stereotypes?" *Debates in Digital Humanities* (Minneapolis: University of Minnesota Press, forthcoming 2018). The continuing debate is apparent in a number of the recent considerations of "distant reading" in *PMLA* 132.3 (2017).

study for identity issues and quantitative analysis in general. What follows is a brief overview of the essentialist / constructionist debates about gender identity as recorded in publication events, best accomplished out of chronological order, in order to work through what is at stake in a number of definitional and theoretical matters. The first essay, from 1990, grounds this overview in practical political activities before getting theoretically complicated.

1990

In a collection called *Conflicts in Feminism*, Ann Snitow described a “divide” between “equality” and “difference” feminism.¹⁰ Snitow describes the conflict between feminists who see women as equal to men, on the one hand, and, on the other, feminists who see women as fundamentally different than men. For equality feminists, “sex hierarchy [is] social, not natural” (Snitow 28). Although Snitow recognized that “difference feminism” could involve theorizing essential differences between men and women, or simply cultural differences, she insisted that “a common divide keeps forming in both feminist thought and action between the need to build the identity ‘woman’ and give it solid political meaning and the need to tear down the very category ‘woman’ and dismantle its all-too-solid history” (1). As an activist, Snitow’s “Gender Diary” records the necessity of switching back and forth between equality and difference to achieve various political goals, but this paradox of needing both to retain and dispense with identity categories dogs theory as well as practice.

1985

Toril Moi opened *Sexual/Textual Politics* by faulting the liberal humanist readings of feminists looking to find “women’s experience” in texts. Writing five years before Snitow, Moi proposed a taxonomy that anticipates Snitow’s in some respects, helping to set the terms for the debate. She contrasts “equality” or “liberal feminism” with “difference feminism” in which “Women reject the male symbolic order in the name of difference. Radical feminism. Femininity extolled.” But

¹⁰ Ann Snitow, “A Gender Diary,” in *Conflicts in Feminism*, ed. Marianne Hirsch, Evelyn Fox Keller (New York: Routledge, 1990), 9–43.

Moi's model incorporated a third position via Julia Kristeva, bringing French feminism and deconstructive modes of reading into the picture: "Women reject the dichotomy between masculine and feminine as metaphysical."¹¹ Binary oppositions generate meaning through difference from each other, not through reference to physical reality, and so in that sense, the gender dichotomy is "metaphysical," but she also means "metaphysical" in the sense of providing presence and origin (Moi 119). The invocation of "difference" in feminist theory influenced by European philosophy draws on structuralist linguistic and anthropological insights that meaning is generated through difference, for instance that "woman" is not "man," and Derridean notions of "différance" or the deferral of meaning: both resonances stressed a lack of intrinsic essence or truth. Moi describes in detail the feminism of Hélène Cixous as combining a notion of diacritical "difference" with the Derridean "différance" which goes beyond merely the "critique of binary logic" to the critique of presence in written (and even spoken) words: language does not provide presence but defers it (Moi 104-109); to repeat a famous axiom of French feminism, "The woman does not exist."¹²

1989

In her book *Essentially Speaking* (1989), Diana Fuss laid out the "essence versus difference" theoretical models for understanding identity that proliferated during the 1980s: "essentialism is typically defined in opposition to difference"; it is "a complex system of cultural, social, psychological, and historical differences, and not a set of pre-existent human essences, [that] position and constitute the subject" (xii). Here, "difference" is used to mean that identities are generated only through the Western European white male hegemonic activity of differentiating self from Other as formulated variously by Hegel, Freud, Sartre,¹³ Beauvoir, Lacan,¹⁴ and Derrida.¹⁵

In her first chapter, Fuss classifies identity theorists as either essentialists or constructionists: "while the essentialist holds that the natural is *repressed* by the so-

¹¹Toril Moi, *Sexual / Textual Politics: Feminist Literary Theory* (New York: Methuen, 1985), 12.

¹²From Juliet Mitchell, Jacqueline Rose, ed. and introduction, *Feminine Sexuality*, eds. Juliet Mitchell and Jacqueline Rose, trans. Juliet Mitchell (New York: Norton, 1982), 72.

¹³Linda Martín Alcoff, "Who's Afraid of Identity Politics?" in Moya, Hames-García, eds., 312-344, the section title "A Genealogy of the Problem," 325-334.

¹⁴Stephen Heath, "Difference," *Screen* 19.3 (1978): 51-112; Alcoff, 329; Fuss, 6-12.

¹⁵Fuss, *Essentially Speaking*, 12-19.

cial, the constructionist maintains that the natural is *produced* by the social”¹⁶. She articulates the conflict between these two camps in the chapter called “Identity Politics”:

Is politics based on identity, or is identity based on politics? Is identity a natural, political, historical, psychical, or linguistic construct? What implications does the deconstruction of “identity” have for those who espouse an identity politics? Can feminist, gay, or lesbian subjects afford to dispense with the notion of unified, stable identities or must we begin to base our politics on something other than identity? What, in other words, is the politics of “identity politics”? (Fuss, *Essentially Speaking*, 100.)

In *Concepts of the Self*, Anthony Elliott quotes this passage and argues that Fuss’s critique “shaped the politics of subversion advocated by queer theorists.”¹⁷ The group “Queer Nation” emerged alongside the gay liberation movement as an alternative political mode of activism that did not require investing in “identity.”¹⁸ As Michael Warner puts it in his introduction to a special issue of *Social Text* published in 1991, “Fear of a Queer Planet,” queer politics in contrast to gay identity politics “partially disarticulat[es] itself from other kinds of identity politics, and partially from the frame of identity politics itself.”¹⁹ Lesbian and gay assertions of a distinct and different identity participate in the social marginalization against which they protest: it allows dispensing with homoerotic desire by designating it as belonging to someone else “over there” (and doing so with violence, in the case of projection).²⁰ In contrast, “queer” takes on global proportions insofar as it names not a separate group but something uncannily ordinary. Rather than protesting against historical marginalization as lesbian and gay activists have done, and in the process enacting what Wendy Brown calls a “wounded attachment” to the past,²¹ queer political action performs queer centrality to the nation or the globe (Abelove, *Deep Gossip* 40, 45-46). Queer activism and theory attempt to “dispense with the notion of unified, stable identities” in order to avoid self-

¹⁶Fuss, *Essentially Speaking*, 3.

¹⁷Anthony Elliott, *Concepts of the Self* (Malden, MA: Polity P, 2008), 132.

¹⁸Henry Abelove, “From Thoreau to Queer Politics,” *Yale Journal of Criticism* 6.2 (1993): 17-27. Rpt. in *Deep Gossip* (Minneapolis: University of Minnesota Press, 2003), 29-41. For use of the term “queer,” see Lisa Duggan, “Making it Perfectly Queer,” *Socialist Review* 22.1 (1992): 11-31. See also David M. Halperin, “The Normalization of Queer Theory,” *Journal of Homosexuality* 45.2-4 (2003): 339-343.

¹⁹Michael Warner, “Introduction: Fear of a Queer Planet,” *Social Text* 29 (1991): 3-17, 11.

²⁰Possibly the first article to express concerns about the danger of demarcating “gay identity” is Henry Abelove’s “Freud, Male Homosexuality, and the Americas,” first published in 1985-6, and subsequently reprinted in *Deep Gossip*, 1-20.

²¹Wendy Brown, *States of Injury: Power and Freedom in Late Modernity* (Princeton, NY: Princeton UP, 1995), 66-76, cited in Moya, Introduction, 8.

marginalization. By insisting that queerness is essential to American identity, not peripheral to it, queer theory calls into question the value of defining any identity at all.

1984

In a much cited and reprinted interview first published in an Australian journal, Gayatri Spivak admonished her “positivist feminist colleagues,” warning them “that essentialism is a trap” too easy to fall into when theorizing identity. Work that is concerned with retrieving “the voice” of non-elite colonized subjects, she argued, needs to be read “from within but against the grain [T]he project to retrieve the subaltern consciousness [should be read as] the attempt to undo a massive historiographic metalepsis and ‘situate’ the effect of the subject as subaltern. I would read it, then, as a strategic use of positivist essentialism in a scrupulously visible political interest.”²²

The goal is to acknowledge the subjective effects of belonging to an identity constituted historically through oppression without believing that the identity itself exists independently of those historical conditions. The phrase “strategic essentialism” took off in a way that Spivak herself had not anticipated. She defines strategic essentialism most clearly in a later interview as “the strategic use of an essence as a mobilizing slogan or masterword like woman, or worker, or the name of any nation that you would like” Spivak adds that she regrets having launched the concept since it is so often used as an “alibi”: she would rather that cultural theorists investigate “how ourselves and others are . . . essentialist, without claiming a counter-essence disguised under the alibi of a strategy.”²³ That is, cultural theorists analyzing identity participate in essentializing, no matter what disclaimers introduce their work.

Three features of Spivak’s intervention are important here, both to subsequent debates within cultural studies and to our concern with quantitative analysis: first, the connection she makes between identity essentialists and positivists; second,

²²Gayatri Spivak, *In Other Worlds: Essays in Cultural Politics* (New York: Methuen, 1987), 205. Rpt. of an earlier interview with Elizabeth Grosz, “Criticism, Feminism and the Institution,” *Thesis Eleven* 10/11 (1984-85): 175-88. (Rpt. also Gayatri Spivak, *Postcolonial Critic*, ed. Sarah Harasym [New York: Routledge, 1990], 1-16).

²³“In a Word,” interview with Ellen Rooney, 1988. In *The Essential Difference*, ed. Naomi Schor and Elizabeth Weed (Bloomington: Indiana UP, 1994), 151-184. Rpt. in *Contemporary Literary Criticism*, ed. Jeffrey W. Hunter, vol. 233 (Detroit: Gale, 2007).

the idea that the essentialism of identity categories cannot just be forsworn in any simple way by a cultural theorist; and third, the problem of history. To elaborate on the third: how can a cultural critic counteract the elision of non-dominant histories, history as the history of the winners (“massive historiographic metalepsis”), without essentializing identities? How does one study the history of woman, or even women, without using the category of woman to mean something consistent through time (s.a. Fuss 3-4)?

It might seem as though categorizing “man” and “woman” for the sake of quantitative analysis *necessarily* participates in a kind of “naïve empiricism” according to which such categories are self-evidently referential.²⁴ Joan Scott explicitly decries numerical analyses in “Gender: A Useful Category of Historical Analysis” (1986)²⁵ accepting M/F as quantities (1 and 2) would mean accepting the idea that gender can be distinguished based on “the single variable of physical difference”²⁶; it would undermine feminism’s necessary “refusal of the fixed and permanent quality of the binary opposition, a genuine historicization and deconstruction of the terms of sexual difference”²⁷. However, one way to analyze the various attributes accorded to a gender that is reconstructed throughout history is to search through large amounts of data using that very binary category—that is, investigating “the origins and consequences” of the social category of the gender binary and its surrounding practices.²⁸ Deborah Verhoeven has commented that Scott’s notion of gender as an analytic category paves the way for scaling up historical inquiries.²⁹ In “A Sociology of Quantification,” Wendy Espeland and Mitchell Stevens argue that quantification is a mode of “commensuration” that “creates a special type of relationship among objects”: it “unites objects by encompassing them under a shared cognitive system.”³⁰ The process of commen-

²⁴“The desire for ‘truth’ or ‘objective’ knowledge” about identities, Paula Moya writes, is seen by postmodernists “as resting on a naively representational theory of language that relies on the following mistaken assumptions: first, that there is a one-to-one correspondence between signs and their extralinguistic real-world referents; and second, that some kind of intrinsic meaning dwells in those real-world referents, independent of human thought or action.” (Moya, Introduction, 5).

²⁵Joan W. Scott, “Gender: A Useful Category of Historical Analysis,” *The American Historical Review* 91.5 (1986): 1053-1075.

²⁶Scott, “Gender,” 1059.

²⁷Scott, “Gender,” 1065.

²⁸Geoffrey C. Bowker, Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences* (Cambridge, MA: MIT Press, 1999), 5. Bowker and Star see this work as an archeology in the Foucauldian sense of the word.

²⁹Discussion preceding a panel at the Digital Humanities 2016 conference: see Brown, S., Clement, T., Mandell, L., Verhoeven, D., Wernimont, J., “Creating Feminist Infrastructure in the Digital Humanities.” In *Digital Humanities 2016: Conference Abstracts*. Jagiellonian University & Pedagogical University, Kraków (2016), 47-50.

³⁰Wendy Espeland, Mitchell Stevens, “A Sociology of Quantification,” *European Journal of Sociology* 49.3 (2008): 401-36; 408.

uration requires first creating the “objects” to be classified, and insofar as quantitative analysis can be used to find historically situated attributes of social classifications, it can give us a clue about “shared cognitive systems” of the past. As Donna Haraway puts it, “‘Gender’ was developed as a category to explore what counts as a ‘woman,’ to problematize the previously taken for granted.”³¹ Gender analysis—determining what has been counted as feminine and masculine through time—can go hand in hand with quantitative analysis.

That cultural analytics can historicize gender can be seen in this issue, for instance in the article by Ted Underwood, David Bamman, and Sabrina Lee. They track sorting according to binary gender categories in novels published over 170 years, arguing that, as time goes on, characters’ actions and attributes are less strongly marked by gender. And, in the case of racial and ethnic identities, Mark Algee-Hewitt, J.D. Porter, and Hannah Walser track the migration of identifying features from one identity to another, as well as those that “stick,” in novels published over the course of 130 years.

Critics have suggested that the theoretical sophistication of these essays may not be typical in the field of Digital Humanities (DH). Miriam Posner argues that the radical potential of DH is unrealized precisely because “most of the data and data models we’ve inherited deal with structures of power, like gender and race, with a crudeness that would never pass muster in a peer-reviewed humanities publication.”³² She highlights the crudeness of census data using Martin Schoeller’s *National Geographic* project “The Changing Faces of America” in which clicking on a photographic image of a person leads to a juxtaposition of the complex and often multiple categories with which they identify themselves, and the census box that they check.³³ Nevertheless one can find more nuanced use of models within fields such as the natural and social sciences where statistics have been absorbed in a range of methodologies. As Katherine Bode puts it,

While [the humanist’s] sense of numbers as an imperfect and mediated representation might not be the exact way they are discussed in the sciences, no scientist approaches statistics as neutral, true and infallible. Awareness of the way scientists interrogate - rather than simply accept or promote - statistical measures is often lacking in current humanities’ debate about quantitative approaches and their

³¹ Donna Haraway, “Gender for a Marxist Dictionary: the Sexual Politics of a Word,” in *Simians, Cyborgs, and Women: the Reinvention of Nature*. London: Free Association Books, 1991. Rpt. in *Women, Gender, Religion: a Reader*, ed. Elizabeth A. Castelli (New York: Palgrave, 2001), 49-75, 67.

³² Miriam Posner, “What’s Next: The Radical, Unrealized Potential of Digital Humanities,” *Miriam Posner* [blog], 27 (2015). See also James English, “Everywhere and Nowhere: The Sociology of Literature After ‘the Sociology of Literature,’” *New Literary History* 41.2 (2010): v- xxiii.

³³ Martin Schoeller, “The Changing Faces of America,” *National Geographic*, October 2013.

ideological resonances. (Bode, *Reading by Numbers*, 12-13.)

Statisticians well know that their results reflect categorization effects, a difficulty compounded when the categories in question are identities. The apparently incidental effects of the procedures of quantitative analysis must be subjected to critical scrutiny, as in an essay included in this issue: Richard Jean So, Hoyt Long, Yuancheng Zhu interrogate sequence alignment tools derived from genomics and bioinformatics in the course of investigating how novels articulate racial identities and relations. What emerges from that two-pronged inquiry is both the limits of the algorithm from a Critical Race Studies perspective, and the extent to which normative literary history has also tended to reify racial differences.

1991

Further challenges to simplistic models of identity have emerged from the study of intersectionality. Emanating from a landmark essay by Kimberlé Crenshaw, “Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color,” that appeared in the *Stanford Law Review* in 1991,³⁴ the “intersectionality theory” promoted by mixed-methods sociologists faults quantitative analyses that organize data at the outset according to discrete identity categories without investigating the ways in which they overlap and inflect each other. Intersectionality, Erez Levon tells us, is grounded in “the belief that no one category (e.g., ‘woman’ or ‘lesbian’) is sufficient to account for individual experience or behavior.”³⁵ While research in intersectionality often requires a mixed methods approach (qualitative and quantitative), cultural analytics can be used to investigate multiple identity categories. In this issue, Elizabeth Evans and Matthew Wilkens examine how ethnicity and national origin are intertwined with genre and geography. Additionally, insofar as they prepare the way for identities to be multiplied using their data and methods, Evans and Wilkens may be seen to engage in intersectional inquiry.

³⁴ Kimberlé Crenshaw, *Stanford Law Review*, 43.6 (1991): 1241-1299.

³⁵ Erez Levon, “Integrating Intersectionality in Language, Gender, and Sexuality Research,” *Language and Linguistics Compass* 9.7 (2015): 295-308, 295.

1988

An interesting feature of quantitative versus qualitative debates in sociology is the emphasis among mixed-methods researchers on “giving voice” to the subjects of their studies.³⁶ Spivak’s landmark essay, “Can the Subaltern Speak?” (1988) delineates how difficult a project that might be, perhaps above and beyond what most sociologists might assume.³⁷ But insofar as their methods involve actually speaking with the subjects of sociological analysis, the project is not possible in any obvious way for literary historians. There are, however, things that can be done (and here we circle back to the third issue raised by Spivak in her 1985 discussion of “strategic essentialism” and elaborated in “Can the Subaltern Speak?”) to avoid the elision of non-dominant histories. Among the articles in this Identity issue of *Cultural Analytics*, several might be said to open up alternative histories of historically marginalized groups. Evans and Wilkins demonstrate that outsider novelists, foreign to British soil, engaged in colonial and anti-colonial critique of modernist internationalism through the greater use, and greater abstraction, of geographical terminology. Lynne Tatlock, Matt Erlin, Douglas Knox, and Stephen Pentecost demonstrate among readers from the U.S. midwest a resistance to reading only those novels that target specific genders. And Kraicer and Piper demonstrate in recent Anglophone novels the under-representation of women characters and the privileging of heteronormativity.

2000

Paula Moya and Michael Hamer-García’s 2000 anthology *Reclaiming Identity* (see footnote 5) provides a treasure-trove of methodological resources and examples for investigating identities with the “post-positivist realist” approach pioneered by Satya Mohanty.³⁸ In that anthology, Linda Alcoff argues that identity studies is only positivistic insofar as its insights are claimed to be more than provisional:

³⁶For an overview, see Yvonne Lincoln, Susan Lynham, Egon Guba, “Paradigmatic Controversies, Contradictions, and Emerging Confluences, Revisited,” in *The Sage Handbook of Qualitative Research*, 5th edition (Thousand Oaks, CA: Sage Publishing, 2017): 108-150; 124, 130, 132.

³⁷Gayatri Spivak, “Can the Subaltern Speak?” originally published in *Marxism and the Interpretation of Culture*, ed. Cary Nelson, Lawrence Grossberg (Urbana: University of Illinois Press, 1988), 67-111.

³⁸Satya P. Mohanty, *Literary Theory and the Claims of History: Postmodernism, Objectivity, Multicultural Politics* (Ithaca: Cornell University Press, 1997); “Colonial Legacies, Multicultural Futures: Relativism, Objectivity, and the Challenge of Otherness,” *PMLA* 110.1 (1995): 108-18.

Ontologies can be thought of as models of reality useful in science (or in social theory) that approximate the world as it is, thus capturing some truth about it, without enjoying a one-to-one correspondence with categories of entities as they exist completely independently of human languages or human practices. (Alcoff, “Who’s Afraid of Identity Politics?” 316.)

The NovelTM group therefore offers these investigations as provisional, expandable, re-workable with modified or other categories, taxonomies, and ontologies.

2018

Despite the risks, there is something to be said about the particular value of quantitative analysis for retrieving marginalized or silenced histories as well as understanding their continued marginalization into the present. In *Deep Gossip* (2003), Henry Abelove describes the continuity between gay liberation histories and queer politics: if the latter is intent upon destabilizing identity categories, he argues, some of the historical work on gay identity in fact leads the way: it is “just a step,” he argues, “from historicizing to destabilizing.”³⁹ One particular feature of doing history at scale is that it allows taking more and more ephemera, into account. As Bode puts it,

quantitative methods allow us to explore aspects of the literary field, especially trends and patterns, broad developments and directions, that would otherwise remain unrepresented and unrepresentable. (Bode, *Reading by Numbers*, 13.)

It may be that numbers, if understood as not transparently readable, can provide another method for “giving voice.” The NovelTM group has embraced the challenge to devise better methods for addressing social identities and power dynamics. So, for example, in “Crossing Over: Gendered Reading Formations at the Muncie Public Library, 1891-1902,” Tatlock et al. combine text mining with borrower records and a history of institutions (book market and libraries) in order to provide a dialectical account of the social and institutional forces that push readers toward gender normative texts as well as how readers push back through book selection.

³⁹ Abelove, *Deep Gossip*, 54.

Historicizing helps to destabilize identities, and cultural analytics can make visible a kind of history that we have never seen before. The analysis of literary texts, whether as objects of consumption or through their textuality, introduces another fruitful layer of complexity that stresses the extent to which identity is always already mediated. In the case of literary investigation, quantitative analysis can engage with and unpack the discursive construction of identities in novels.

Three of the essays in this issue, all mentioned briefly above, deal with racial and ethnic identities. In “Computational Method and the Critique of Race: Racial Difference and the US Novel at Scale, 1880-2000,” Richard Jean So, Hoyt Long, and Yuancheng Zhu take up the call, made recently by Roopika Risam, Kim Gallon, Amy Earhart, Lauren Klein, and others, to develop a form of computational criticism and distant reading that is commensurable with the methods and arguments of Critical Race Studies.⁴⁰ The essay takes as its assumption that scientific and quantitative methods often reify race and support racial stratification (eugenics, the Bell Curve, etc.). Thus, any “big data” or computational method applied to racial minority authors and texts is in great peril of simply reproducing reified and problematic views of racial identity. So, Long, and Zhu argue that canonical methods in textual analytics—specifically, sequence alignment—can be productively deformed and reconstructed through an attention to the critique of race not only to produce new “large scale” views of cultural and literary history, but also to advance the work of critique by challenging the assumptions of the algorithm itself. They combine theory and close reading with sequence alignment analysis to critique the racial homogeneity and universalism, nominally supported by the method of sequence alignment, through a case study focused on the modern American novel and the question of racial difference. A new story about racial difference and the US novel emerges from testing the limits of the algorithm, which are often the limits of normative literary history itself. That is, a creative use of computation, animated by critique, draws attention to what has been written out of literary history while at the same time advancing the work of racial critique.

In “Representing Race and Ethnicity in American Fiction: 1789-1920,” Mark Algee-Hewitt, J.D. Porter, and Hannah Walser explore the relationship between

⁴⁰Roopika Risam, “Beyond the Margins: Intersectionality and the Digital Humanities,” *DHQ* 9.2 (2015), and “Navigating the Global Digital Humanities: Insights from Black Feminism,” in *Debates in Digital Humanities*, ed. Matthew K. Gold, Lauren Klein (Minneapolis: University of Minnesota Press, 2016); Kim Gallon, “Making a Case for the Black Digital Humanities,” in *Debates in Digital Humanities*, ed. Matthew K. Gold, Lauren Klein (Minneapolis: University of Minnesota Press, 2016); Amy Earhart, “Can Information be Unfettered? Race and the New Digital Humanities Canon,” in *Debates in Digital Humanities*, ed. Matthew K. Gold, Lauren Klein (Minneapolis: University of Minnesota Press, 2012), 309-318; Lauren Klein, “The Image of Absence: Archival Science, Data Visualization, and James Hemings,” *American Literature* 85.4 (2013): 661-688.

the novel and the discourses of identity that have shaped the understanding of ethnicity, race and ancestry in America. The authors use a combination of collocate analysis and word embeddings on a corpus of over 18,300 novels written between 1789 and 1920 to identify the terms that cluster with distinct markers of racial and ethnic identity over time. Through these methods, they are able to assess the ‘stickiness’ of identity markers: how certain terms remained clustered with particular identities and how others moved in concert between subsequent waves of immigrant groups. Their project not only permits the exploration of how and when certain discourses of identity solidified in the imagination of the reading public, but also how literature itself shaped the American concepts of race and ethnicity. This article deploys race as an analytic category, showing identity categories in the making as races and ethnicities are discursively endowed with specific attributes. Algee-Hewitt, Porter, and Walser thus historicize race and ethnicity in demonstrating the extent to which identities are continually shifting and trading their discursive markers.

Elizabeth Evans and Matthew Wilkens’s contribution, “Nation, Ethnicity, and the Geography of British Fiction, 1880-1940,” draws on four distinct, substantial collections of British literature published in the late nineteenth and early twentieth centuries to assess the applicability of existing critical claims about modernist internationalism to large runs of the period’s fiction and trace the complex interplay of insider and outsider status in writing by authors born outside the UK. They find that, within a framework of broad continuity, the decades leading up to the Second World War saw increasing literary attention to locations beyond Britain’s borders, but there was a drop in the overall rate at which named locations appeared, shifts they attribute to forces cultural, political, and aesthetic. Evans and Wilkens show that foreign writers favored an especially geographically intensive style that drew more heavily on abstractions than on the details of setting, which, they argue, was linked to such authors’ greater engagement with the era’s colonial and anticolonial politics. Beyond its specific inquiries, the article describes a set of broadly applicable methods for analyzing textual geography at scale and provides a rich dataset for other scholars studying what the authors call the “long modernist era.”

Two articles in this special issue emphasize that quantitative methods can be used as, or with, linguistic and literary analysis, mirroring the mixed methods approach advocated by social scientists. In their contribution, “Self-Repetition and East Asian Literary Modernity, 1900-1930,” Hoyt Long, Anatoly Detwyler, and Yuancheng Zhu approach identity as it relates to the construction of a modern narrative voice characterized by psychic interiority and a vernacular style. Specifically, they analyze linguistic redundancy as a constitutive feature of liter-

ary self-identity in early 20th-century Japanese and Chinese fiction. Using measures of information and diversity in diction combined with other simple textual features, they find a tendency toward repetition in self-referential or psychologically oriented narratives and argue for its role as conscious stylistic innovation within East Asian literary modernity. This finding, which holds across distinct but interrelated national literary contexts, shows how quantitative methods can contribute to translanguing comparison. It also offers an opportunity to reflect on the longer history of linguistic measures of redundancy, in particular their origins in mid-century psycholinguistics as indexes of mental aberration. Long, Detwyler, and Zhu's approach to history is a reminder of how intertwined is the relation of measurement to knowledge of the psychological self, and of how embedded this relation is in ideas about language.

"Crossing Over: Gendered Reading Formations at the Muncie Public Library, 1891-1902," by Lynne Tatlock, Matt Erlin, Douglas Knox, and Stephen Pentecost interweaves thick description (literary analysis) with a consideration of broad social patterns, providing a finely-grained consideration of gender-normative reading in nineteenth-century Muncie, Indiana. They examine intra- and extra-textual pressures that impinged upon reader choice in the period in order to discuss "crossover" and "convergent" reading practices. "Crossover reading" is the reading performed by the non-targeted gender of overtly gender-targeted fiction that was otherwise most frequently read, as intended, by persons of the targeted gender. "Convergent reading" refers to the selection of fiction—not overtly specified to gender, e.g., as "books for women" or "books for men"—by persons of either gender at similar rates relative to the respective gender group. The traditional analytic categories of gender-convergent and gender-crossover reading actually break down in the case of novels by Horatio Alger that combine elements typical of books geared toward one gender of reader or another. This insight only emerges as a result of placing aggregate borrowing patterns in dialogue with text-mining techniques designed to find novelistic features typed by gender, analyzing simultaneously literary qualities and statistical patterns. Additionally, Tatlock, Erlin, Knox, and Pentecost, offer a situated understanding of gender; their article performs gender trouble through its investigation of crossover reading and the crossover textuality of Horatio Alger.

Two other essays in this issue also deal with gender as an analytic category. Relatively fixed and fluid forms of identity rub shoulders in an essay about gender by Ted Underwood, David Bamman, and Sabrina Lee. In a familiar, social-scientific mode, this essay tells a story about the shifting demographic balance of English-language fiction. Peering through this lens, signs of progress are hard to discern. Men increasingly outnumber women in the period 1850-1970, whether we con-

sider the grammatical gender of characters or the names of authors. Changes at the end of the twentieth century only seem to restore a nineteenth-century status quo. But this essay also uses machine learning to consider shifting perspectives on gender itself. A perspectival approach to the topic reveals a more progressive arc: as we move from the nineteenth century through the twentieth, the actions and attributes of characters are less strongly sorted into binary gender categories. Perspectival analysis even reveals how the concepts of gender that organize characterization may be inflected by the gender identification of the author. As mentioned above, this article demonstrates the changing construction of gender through time.

Eve Kraicer and Andrew Piper's essay, "Socializations: Gender, Genre, and the Social Networks of the Contemporary Novel," begins in many ways where Underwood et al. leave off. Their dataset starts after 2005, allowing insights into the gendered nature of character in the contemporary novel. At the same time, Kraicer and Piper's focus on the hierarchical positioning of gender within the novel's social networks offers further insight into the social reality presented within novels, what they identify as a key form of "socialization." Where Underwood et al. identify the declining semantic differentiation between men and women in novels, Kraicer and Piper's work highlights the way such de-gendering of the novel's diegetic universe is offset by the persistence of very clear gender hierarchies. Using a collection of ca. 1,300 contemporary novels in English drawn from seven different genres, Kraicer and Piper show how recent anglophone novels uniformly under-represent women characters, regardless of either genre or author gender. Additionally, these novels manifest a decided orientation towards depicting heterosexual relationships. Taken together, we see two forms of patriarchy being expressed in the English language novel today: the relative absence of one gender and a heteronormative sociability on the other. These insights hold across a broad range of novels from popular genre fiction like romance or mysteries to more up-market works that have received literary prizes or been reviewed in the New York Times.

For Kraicer and Piper, computational methods are important because they allow for a larger-scale understanding of the way gendered identities are not only constructed within contemporary fiction but also socially positioned. This social organization of gender conveys powerful facts about how readers are themselves "socialized" through contemporary novel reading. They see their work as a direct response to calls by feminist scholars to draw attention to "processes and possibilities of social and cultural transformation."⁴¹ Knowledge of such hierarchies

⁴¹ Alexis Lothian and Amanda Phillips, "Can Digital Humanities Mean Transformative Critique?" *E-Media Studies* 3.1 (2013): DOI:10.1349/PS1.1938-6060.A.425

of (in)visibility provide in their view an important basis for a form of cultural advocacy that would seek to undo the fictional realities that potentially inform everyday norms.⁴²

All of these essays thus represent contributions to their specific fields of inquiry, but also a rich set of efforts to advance our understanding of what it means to mobilize identity categories within large-scale computational analysis of literary phenomena. The data modelling this work requires necessarily engages in both abstraction and reduction, but the very act of modelling carries with it the seeds of a constructionist recognition that a phenomenon could be modelled differently, and as a number of the essays show, conjoining diverse categories or pluralizing the modes of inquiry can reveal the dynamic and contingent nature of identity categories. At the same time, however, and as the debates surveyed all too briefly above indicate, these categories are easily reified because they readily map onto categories that have been and still often are considered fixed and essential, and that do real, politically charged work in the world.

This debate is increasingly pressing of late, with a rise in critical attention being given to methodology within the digital humanities more generally. Alan Liu considers ours an infrastructure moment, and his call for “critical infrastructure studies”⁴³ has further focused the vigorous discussion of the underlying conditions of digital humanities practices. This includes heightened scrutiny of the politics of digital humanities infrastructures ranging from the institutional ones that embed and perpetuate structural inequalities to the tools and techniques employed in the kind of work promoted by *Cultural Analytics* which, as Lauren Klein has observed, “also derive from flaws in cultural and conceptual structures” of the kind we have explored above at length in the conundrum of how to mobilize categories related to gender.⁴⁴ Catherine D’Ignazio and Klein, working from a feminist ethical and theoretical position, for instance, call for a very high standard for data visualization that not only takes on the conceptual matter of binary categories on which we have focused here, but also brings in other important considerations including perspective, embodiment and affect, and transparency with respect to provenance and labour associated with datasets.⁴⁵ The commitment of the NovelTM group to transparency with regard to the provenance of datasets,

⁴²For a discussion of the way “cultural processes” mediate between large-scale social effects and individual cognitive processes that rely on schemas and tokenization, see Michèle Lamont, Stefan Beljean, and Matthew Clair, “What is Missing? Cultural Processes and the Making of Inequality,” *Socio-Economic Review*, 12.3 (2014): 573–608.

⁴³Alan Liu, “Drafts for Against the Cultural Singularity (book in progress),” Alan Liu [blog] 2 May 2016.

⁴⁴Lauren Klein, “Distant Reading after Moretti,” *Lauren F. Klein* [blog] (January 2018).

⁴⁵Catherine D’Ignazio and Lauren F. Klein, “Feminist data visualization,” in Workshop on Visualization for the Digital Humanities, VIS4DH (Baltimore, IEEE, 2016).

and to making those datasets available for the evaluation and reproducibility of the work presented here, as well as the basis for further work, is a contribution over and above that of the particular results. Documenting and archiving datasets is not trivial work, nor is it typically rewarded. The datasets made available in connection with this work contribute to establishing a crucial infrastructure for the investigation of cultural identities using digital humanities methods, including of recent texts that cannot circulate freely.⁴⁶ As Klein argues, “it may very well be that a distant view that is trained on power, and that is self-reflexive about the forces that enable it—cultural and conceptual as well as computational—can contribute significantly to the project of dismantling structural power.”⁴⁷ The contributions clustered here will enable the ongoing debate over how cultural analytics can engage ethically and convincingly with thorny questions of identity to proceed productively on the basis of both theory and practice.

⁴⁶All relevant data and code is available at the *Cultural Analytics* dataverse repository at <https://dataverse.harvard.edu/dataverse.xhtml?alias=culturalanalytics>.

⁴⁷D’Ignazio and Klein, “Feminist data visualization.”